PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
C12N 15/00, 15/85, 15/90 C12N 15/62 // C12N 15/58	A1	WO 91/13151
		(43) International Publication Date: 5 September 1991 (05.09.91)

(21) International Application Number: PCT/US91/01222

(22) International Filing Date: 22 February 1991 (22.02.91)

(30) Priority data:
483,450 22 February 1990 (22.02.90) US

(71) Applicant: BIOGEN, INC. [US/US]; 14 Cambridge Center, Cambridge, MA 02142 (US).

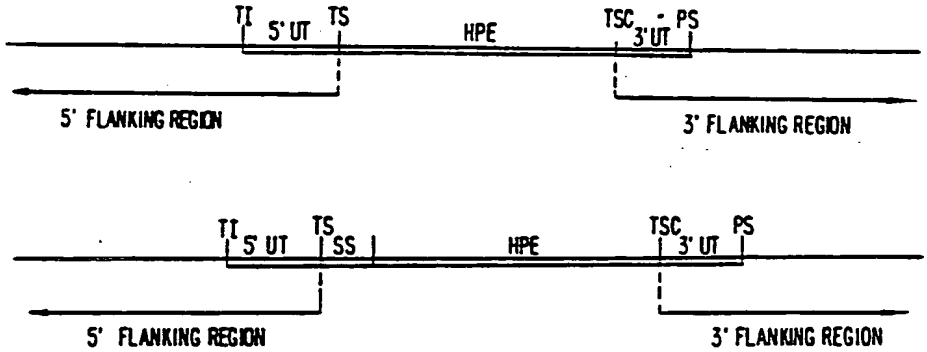
(72) Inventor: MEADE, Harry ; 62 Grasmere Street, Newton, MA 02158 (US).

(74) Agent: HALEY, James, F., Jr.; Fish & Neave, 875 Third Avenue, New York, NY 10022-6250 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CM (OAPI patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IMPROVED EXPRESSION OF POLYPEPTIDES



TI= TRANSCRIPTION INITIATION SITE
UT= UNTRANSLATED REGION
TS= TRANSLATIONAL START SIGNAL
SS= SIGNAL SEQUENCE ENCODING REGION
HPE= HETEROLOGOUS POLYPEPTIDE ENCODING REGION
TSC= TRANSLATION STOP CODON
PS= POLYADENYLATION SIGNAL

(57) Abstract

This invention relates to processes and intermediates for improving the level of production of a desired polypeptide in a recombinant host whose genome has integrated into it an island of expression comprising in the 5' to 3' direction, a 5' flanking region, the heterologous polypeptide encoding sequence and a 3' flanking region. The island of expression of this invention permits the expression of the integrated heterologous DNA sequence to be substantially dependent on its copy number and to be substantially independent of its position of integration in the host genome.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MG**	Madagascar
AU	Australia	**FI**	Finland	**ML**	Mali
BB	Barbados	**FR**	France	**MN**	Mongolia
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Faso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GN**	Guinea	**NL**	Netherlands
BJ	Benin	**GR**	Greece	**NO**	Norway
BR	Brazil	**HU**	Hungary	**PL**	Poland
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		

IMPROVED EXPRESSION OF POLYPEPTIDES

TECHNICAL FIELD OF INVENTION

This invention relates to processes and intermediates for improving the level of production of a desired polypeptide in a recombinant host. More particularly, this invention relates to an "island of expression" -- a segment of DNA which contains a DNA sequence encoding a heterologous polypeptide -- and the use of the island of expression to transfect a host. Hosts harboring this island of expression produce a surprisingly high level of the desired heterologous polypeptide. Incorporation of the island of expression into a host permits the desired heterologous polypeptide to be expressed substantially independent of its position of integration in the host genome and substantially dependent on the number of copies of the island of expression which integrate into the host genome.

BACKGROUND ART

It is well known that polypeptides can be expressed and secreted by hosts transformed or transfected with a DNA sequence coding for that polypeptide. For example, Gilbert et al., United States Patent 4,565,785 (1986) and L. Villa-Komaroff et al., "A Bacterial Clone Synthesizing Proinsulin",

Proc. Natl. Acad. Sci. USA, 75, pp. 3727-31 (1978) have shown that a selected polypeptide can be synthesized within a bacterial host and excreted through the host membrane. A similar process can be carried out in animal cells. J. Doehmer et al., "Introduction Of Rat Growth Hormone Gene Into Mouse Fibroblasts Via A Retroviral DNA Vector: Expression And Regulation", Proc. Natl. Acad. Sci. USA, 79, pp. 2268-72 (1982). Recombinant proteins have even been expressed in mammals through transgenic incorporation of an expression system into the pronucleus of a fertilized embryo. D. Bucchini et al., "Pancreatic Expression Of Human Insulin Gene In Transgenic Mice", Proc. Natl. Acad. Sci. USA, 83, pp. 2511-15 (1986); K. Gordon et al., "Production Of Human Tissue Plasminogen Activator In Transgenic Mouse Milk", Bio/Technology, 5(11), pp. 1183-87 (1987).

However, to date, none of these techniques has been consistently successful in permitting large amounts of a desired heterologous polypeptide to be expressed by a host which has integrated into its genome a heterologous polypeptide encoding sequence. This is particularly surprising in view of the high level of native protein production occasioned from the very same expression control sequences in their native environments. For example, milk specific expression control sequences permit large amounts of native proteins, e.g., casein, to be produced in and secreted from mammary glands. The very same milk specific expression control sequences, however, have not been demonstrated to induce large amounts of heterologous polypeptides when operatively linked to heterologous polypeptide encoding sequences. See, for example, C.W. Pittius et al., "A Milk Protein Gene Promoter Directs The Expression Of Human Tissue Plasminogen Activator

cDNA To The Mammary Gland In Transgenic Mice", Proc. Natl. Acad. Sci. USA, 85, pp. 5874-78 (1988). The level of expression in these latter constructions is also independent of the number of copies of the heterologous polypeptide encoding sequence integrated into the host genome. Furthermore, the level of expression is subject to positional effects, i.e., it is dependent on where the heterologous polypeptide encoding sequence is integrated into the genome. K.F. Lee et al., "Tissue-Specific Expression Of The Rat Beta-Casein Gene In Transgenic Mice", Nucleic Acids Res., 16(3), pp. 1027-41 (1988).

Accordingly, the need exists for a method of increasing the expression of DNA sequence encoding a heterologous protein or polypeptide independent of its site of integration in the host genome. Moreover, such methods should provide expression that is dependent upon the number of copies integrated into the host genome so that expression levels may be controlled.

DISCLOSURE OF THE INVENTION

The present invention solves these problems by providing an "island of expression" containing a DNA sequence which codes for a desired heterologous polypeptide. The island of expression of this invention provides for the first time, high level, position-independent and copy number-dependent expression of a DNA sequence coding for a heterologous polypeptide.

As is depicted in Figure 1, the island of expression of this invention comprises, in the 5' to 3' direction, a 5' flanking region, a heterologous polypeptide encoding sequence (coding for the desired heterologous protein or polypeptide) and a 3' flanking region. The 5' flanking region comprises, in the 5'

and 3' direction, 5' expression control sequences and a 5' untranslated region. The expression control sequences are operatively linked to the heterologous polypeptide encoding sequence. The 5' untranslated region begins at a transcription initiation site and ends at the translational start site of the heterologous polypeptide encoding sequence. The 3' flanking region comprises in the 5' to 3' direction, a 3' untranslated region, and 3' expression control sequences, those control sequences being operatively linked to the heterologous polypeptide encoding sequence. Finally, the 5' and 3' flanking regions of the island of expression invention are characterized by a sufficient size and structure effective to render the level of production of the desired protein or polypeptide substantially dependent on the copy number of the island of expression integrated into the host genome and substantially independent of its integration site.

This invention also relates to the use of the island of expression to transfect a host and to those transfected hosts. Hosts which have integrated the island of expression into their genome produce high levels of the heterologous polypeptide encoded by a DNA sequence within that island of expression. Furthermore, the expression processes of this invention are substantially dependent on the copy number of the island of expression integrated into the host genome and independent of the site of integration, which advantageously allows expression levels to be manipulated.

In a preferred embodiment of this invention, the island of expression also includes a DNA sequence coding for a signal peptide. This signal sequence coding region is fused to, and in reading frame with,

the 5' end of the heterologous polypeptide coding sequence. The signal sequence coding region is also operatively linked to the expression control sequences so as to permit a host whose genome carries this preferred island of expression to produce, secrete, and preferably process, the desired protein or polypeptide from the pre-protein or pre-polypeptide coded for by the combined signal-heterologous polypeptide coding sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts a schematic representation of a typical "island of expression" (A) and a preferred "island of expression" (B) in accordance with this invention.

Figure 2 depicts the construction of a plasmid (CAS1288) containing the 5' and 3' flanking regions of bovine alpha S-1 casein.

Figure 3 depicts the introduction of the urokinase structural gene into CAS1288 to yield CAS1295, the island of expression.

DETAILED DESCRIPTION OF THE INVENTION

In order that the invention herein described may be more fully understood, the following detailed description is set forth.

In this description the following terms are employed:

Expression control sequences -- DNA sequences that control and regulate expression of gene products at both the transcriptional and translational level when operatively linked to a structural gene (DNA coding for a polypeptide). They include the promoter and enhancer regions, ribosome binding sites,

polyadenylation signals and other sequences useful in the expression of genes.

Operatively linked -- the linking of 5' and 3' expression control sequences to a heterologous polypeptide encoding sequence so as to permit the expression control sequences to control and regulate the expression and production of the heterologous polypeptide.

Heterologous polypeptide encoding sequence -- a DNA sequence coding for a desired polypeptide or protein that is inserted into the genome of a host. This DNA sequence codes for a polypeptide which is heterologous to either the host, the flanking sequences or both. The heterologous polypeptide encoding sequence optionally contains its own translational start signal at its 5' end and its own translational stop codon at its 3' end. The heterologous polypeptide encoding sequence may also contain its own signal sequence coding region.

Signal sequence coding region -- a DNA sequence which encodes a sequence of typically hydrophobic amino acids called a signal peptide. The signal peptide allows a polypeptide to which it is attached to cross a biological membrane.

Island of expression -- a DNA construct comprising in the 5' to 3' direction, a 5' flanking region, a heterologous polypeptide encoding sequence and a 3' flanking region. The 5' and 3' flanking regions are of sufficient size and structure to render the level of production of the desired protein or polypeptide substantially dependent on the copy number of the island of expression construct incorporated into the host genome and substantially independent of the position of integration of the island of expression in the host genome.

5' flanking region -- is that part of the island of expression which is 5' to the heterologous polypeptide encoding sequence. It includes, in the 5' to 3' direction, 5' expression control sequences and a 5' untranslated region, the expression control sequences being operatively linked to the heterologous polypeptide encoding sequence. The 5' untranslated region typically extends from a transcription initiation site to the translational start site of the heterologous polypeptide encoding sequence.

3' flanking region -- is that part of the island of expression which is 3' to the heterologous polypeptide encoding sequence. It includes, in the 5' to 3' direction, a 3' untranslated region, and 3' expression control sequences. The 3' flanking region may also include all or a portion of the coding sequence from the structural gene originally associated with the 3' flanking region.

DETAILED DESCRIPTION OF THE INVENTION

Although not wishing to be bound by theory, we believe that the island of expression allows as yet undefined factors within the 5' and 3' flanking regions to operate on the expression control sequences and to permit the heterologous polypeptide encoding sequence to be expressed at higher yields. Expression is also dependent on the number of copies of the island of expression construct incorporated into the host genome, thus allowing the level of polypeptide production to be modulated.

The large 5' and 3' flanking regions of the islands of expression of this invention may also provide a buffer zone so that the expression control sequences are isolated from host expression controls which may be exerted by the surrounding DNA into which

the island of expression has integrated. Therefore, no matter where in the host genome the island of expression integrates, the heterologous polypeptide encoding sequence will be expressed at a high level. It carries its own genomic environment along with it, as an "island of expression".

Although not wishing to be bound by theory, we believe that the majority of regions of DNA which may enhance expression from expression control sequences are found in the 5' and 3' flanking sequences of a given structural gene. Therefore, after isolation of a structural gene with its 5' and 3' flanking regions, the structural gene, in accordance with one embodiment of this invention, may be excised in whole or in part and replaced with any heterologous polypeptide encoding sequence so as to permit expression at a level consistent with that of the original structural gene. Alternatively, the heterologous polypeptide encoding sequence may be inserted at the 5' end of the structural gene without concomitant removal of that gene. In that embodiment, the heterologous polypeptide encoding sequence will also be expressed at a level that is comparable to the expression level of the original structural gene.

Among the expression control sequences useful in the various embodiments of this invention are those which direct expression at high levels in particular types of cells or at particular stages of cell growth or differentiation, or under specific culture conditions. Tissue-specific expression control sequences are preferred in the transgenic hosts of this invention.

If mammalian host cells are utilized, useful expression control sequences may be derived from native sequences encoding a highly expressed product from the

host cell itself, or they may be derived from other eukaryotic genes with high levels of expression, such as β-actin, collagen, myosin, albumin, metallothionein and human growth hormone.

A preferred embodiment of this invention provides for the production of proteins in transgenic mammals. This embodiment preferably uses expression control sequences which control and direct expression of gene products in mammary tissue, such as expression control sequences corresponding to casein promoters and the beta lactoglobulin promoter. The casein promoters may, for example, be selected from an alpha casein promoter, a beta casein promoter or a kappa casein promoter. More preferably, the casein promoter and associated expression control sequences are of bovine origin and most preferably are an alpha S-1 casein promoter and associated expression control sequences.

Expression control sequences may even be derived directly from the cells which are to be used as the host for the island of expression construct. A promoter and associated expression control sequences having the desired level of activity in the host must first be identified. The island of expression must be designed so that each island of expression construct which integrates into the host genome is expressed in a copy number-dependent, position-independent manner.

We describe here a means of identifying expression control sequences, cloning the required flanking regions containing these sequences, adding the heterologous polypeptide encoding sequence, and testing whether the resultant construct is an "island of expression" in accordance with this invention.

The first step is to determine a host and conditions which allow a gene homologous to that host to be expressed at a desired level or at specific

times. In the case of tissue culture, CHO cells growing on the collagen beads found in the VERAX™ system are preferably used.

To isolate the expression control sequences for a homologous gene that is expressed at high levels in host cells under selected conditions, an abundantly expressed RNA species must be identified. This may be achieved by preparing a cDNA library from polyA RNA isolated from a selected host cell under selected conditions of induction and growth. The cDNA library is then screened using a labelled aliquot of the same RNA from which the cDNA library was produced. The most positive signals are indicative of those cDNAs whose RNAs are most abundant in the host cell under the selected conditions of induction and growth. The selected cDNAs may then be used to screen genomic DNA libraries prepared from the selected host cells in order to select genomic DNA sequences that correspond to most abundant RNAs. These genomic sequences, typically in cosmids [T. Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982)], may then be analyzed to determine restriction sites, the amount of flanking sequences in the cosmid and the polypeptide coding regions contained therein.

Alternatively, but less preferably, the expression control sequence may be isolated by screening a host cell grown under selected conditions and induction for an abundantly produced protein or polypeptide. This is achieved by analyzing the total polypeptides produced from the host using either SDS polyacrylamide gel electrophoresis (SDS PAGE) or two-dimensional gel electrophoresis. The most abundant polypeptides are identified by the strongest band in an SDS-PAGE gel or the largest spot in a two-dimensional

gel. Once identified, the band or spot is excised from the gel, eluted, and subjected to automated protein sequencing. Oligonucleotides based upon the amino acid sequence obtained from the protein sequencing are then synthesized. These oligonucleotides can then be labeled and used as probes to identify their corresponding genomic sequences from a cosmid library constructed from host cell DNA.

Once a sufficiently detailed restriction map of this abundantly expressed gene has been determined, the coding sequences and intervening sequences of the structural gene may be removed from the cosmids, for example, with appropriate restriction enzymes and replaced with the heterologous polypeptide encoding sequence. Alternatively, the heterologous polypeptide encoding sequence may be inserted 5' to the structural gene. In this embodiment, the structural gene need not be excised. According to a preferred embodiment, the heterologous polypeptide is urokinase, the DNA sequence of which has been isolated and cloned from a genomic library using published sequences as probes. A. Riccio et al., "The Human Urokinase-Plasminogen Activator Gene And Its Promoter", Nucleic Acid Res., 13(8), pp. 2759-71 (1985).

The resulting construct has the DNA sequence coding for the heterologous polypeptide flanked on both sides by the genomic sequences of the abundantly expressed gene which was originally isolated from the host cells. Constructs containing the various lengths of 5' and 3' flanking sequences must be tested to determine what size flanking regions are necessary to direct expression of the heterologous polypeptide encoding sequence in a copy number-dependent, position-independent manner.

To determine that the isolated cosmid contains sufficient 5' and 3' flanking regions to permit an inserted heterologous polypeptide encoding sequence to be expressed at substantially the same level as that of the highly expressed homologous DNA sequence, the selected DNA sequence is transfected into cells in tissue culture or introduced into the genome of an embryos to produce transgenic animals. Preferably, the cells or embryo that will be used for ultimate production are employed in this step. The transformed hosts are then tested for the expression of the heterologous protein by any of a number of well-known assays. These include, but are not limited to, radioimmunoassay, ELISA, immunoblotting and assays which measure the activity of the desired polypeptide. Alternatively and preferably, mRNA levels under a variety of growth conditions are used. This may be achieved by the Northern blot technique using the previously described oligonucleotides (corresponding to the polypeptide sequences) or the cDNAs identified previously as probes.

Because the expression control sequences selected from the host cells demonstrate the ability to direct expression of the homologous gene at a high level under known conditions (e.g., CHO cells growing on collagen beads in the VERAX™ system), it is expected that substantially the same level of expression of the heterologous polypeptide would be seen under those same conditions. Should the cosmid derived DNA sequence not provide such level of expression, then other cosmids containing different lengths of 5' and 3' flanking regions should be analyzed in substantially the same way until an appropriate DNA sequence is located.

The levels of production of the heterologous protein adduced by this sequence are then compared to

the copy numbers of the integrated island of expression. Copy number is determined by appropriate restriction enzyme analysis. The expression constructs which show position-independent, copy number-dependent expression are the optimal "islands of expression" in accordance with this invention.

According to a preferred embodiment the desired polypeptide is secreted by a host harboring an island of expression of this invention. Secretion of polypeptides is accomplished by fusing a DNA sequence coding for a signal peptide to, and in reading frame with, the DNA encoding the heterologous polypeptide. The size of the signal peptide is not critical for this invention. All that is required is that the signal peptide be of a sufficient size and sequence to effect secretion of the heterologous polypeptide. The signal sequence encoding the signal peptide may be exemplified by signal sequences associated in nature with the expression control sequences, signal sequences associated in nature with the desired heterologous protein or polypeptide, signal sequences which are native to the host, signal sequences which are native to the source of the heterologous polypeptide, signal sequences which are native to the source of the expression control sequences and any other sequences encoding functional signal peptides.

Many of the proteins to be expressed are normally secreted and will have their own signal peptide which should be adequate to direct secretion. In this case, the DNA encoding that signal may be included in the heterologous polypeptide encoding sequence that is inserted into the island of expression. To produce a polypeptide that is not normally secreted, it is possible to use a signal sequence from polypeptides which are normally secreted

from the host cells or from other secreted polypeptides. A preferred embodiment of this invention uses sequences encoding milk-specific signal peptides or other signal peptides useful in the maturation and secretion of protein in mammary tissue. These include the signal sequence from alpha S-1 casein. If the heterologous polypeptide to be expressed is associated in nature with its own signal sequence, the signal sequence associated in nature with the heterologous polypeptide coding sequence is the more preferred signal sequence.

The necessary 5' and 3' flanking regions are characterized by the ability to cause expression from the island of expression construct to be position-independent and copy number-dependent. The length of the flanking sequences is not critical as long as these properties are conferred to the expression construct. The upper size limit is defined by the ease of manipulating the DNA. In the original source of the expression control sequences (in the animal or in the cell line), the expression control sequences are flanked, in theory, by the whole chromosome. Present techniques allow the ready manipulation of 40-50 kb segments of DNA. This requires the use of well-known cosmid technology. There may also be a limit on the size of DNA that can be injected through the needles used in embryo manipulations. The preferred technique is to use as large 5' and 3' flanking regions as possible to insure enough insulating region to confer copy number dependence and position independence.

The coding sequence of the desired heterologous polypeptide can be derived from either cDNA, genomic sequences, synthetic DNA or semisynthetic DNA. Among the polypeptide products which may be produced by the processes of this invention are, for

example, coagulation factors VIII and IX, human or animal serum albumin, tissue plasminogen activator (tPA), urokinase, alpha-1 antitrypsin, animal growth hormones, Mullerian Inhibiting Substance (MIS), cell surface proteins, insulin, interferons, interleukins, milk lipases, antiviral proteins, peptide hormones, immunoglobulins, lipocortins and other heterologous protein products.

The desired heterologous polypeptide may be produced as a fusion protein containing amino acids in addition to those of the desired or native protein. For example, the desired heterologous polypeptide of this invention may be produced as part of a larger heterologous protein or polypeptide in order to stabilize the desired protein or to make its purification easier and/or faster. This may be achieved by inserting the heterologous polypeptide encoding sequence into the island of expression at a position 5' to, and in reading frame with, the structural gene, or portion thereof, which was originally associated with the expression control sequences. It will be obvious that such a construct requires removal of the heterologous polypeptide termination codons prior to insertion into the island of expression.

Alternatively, the fusion protein coding region may be constructed prior to insertion into the island of expression. The fusion protein construct may comprise 2 or more heterologous polypeptide encoding sequences or portions therof, as long as the seqeunces are in the same reading frame. Such constructs may be made using techniques known in the art. The fusion protein may then be cleaved, if desired, and the desired protein isolated. The desired heterologous polypeptide may be produced as a fragment or derivative of the polypeptide that was originally associated with

the expression control sequences. Each of these alternatives is readily produced by merely choosing and/or manipulating the correct DNA sequences. Such manipulations are well known in the art.

The above-described island of expression constructs may be prepared using methods well known in the art. For example, various ligation techniques employing conventional linkers, restriction sites, etc. may be used to good effect. Preferably, the islands of expression of this invention are prepared as part of larger plasmids. Such preparation allows the cloning and selection of the correct constructions in an efficient manner as is well known in the art and permits convenient production of large quantities of the island of expression construct.

The particular plasmid is not critical to the practice of this invention. Rather, any plasmid known in the art to be capable of being replicated, selected for, and carrying large pieces of DNA, would be a suitable vehicle in which to insert the islands of expression of this invention. Most preferably, the islands of expression of this invention are located between convenient restriction sites on the plasmid so that they can be easily isolated from the remaining plasmid sequences for incorporation into the desired host.

The selection of an appropriate host for the island of expression invention is controlled by a number of factors recognized in the art. These include, for example, compatibility with the chosen vector, toxicity of the polypeptide products, ease of recovery of the desired heterologous polypeptide, expression characteristics, special processing requirements of the heterologous polypeptide, biosafety and costs. No absolute choice of host may be made for

a particular desired protein or polypeptide from any of these factors alone. Instead, a balance of these factors must be struck with the realization that not all hosts may be equally effective for expression of a particular heterologous polypeptide.

Useful mammalian host cells may include B and T lymphocytes, leukocytes, fibroblasts, hepatocytes, pancreatic cells and undifferentiated cells. Preferably, immortalized mammalian cell lines would be utilized. For example, useful mammalian cell lines would include 3T3, 3T6, STO, CHO, Ltk$^-$, FT02B, Hep2B, AR42J AND MPC1L. Most preferable mammalian cell lines are CHO, 3T3, and Ltk$^-$.

Embryos from various mammals may be used in this invention to produce transgenic animals. The choice of a host embryo may depend on factors such as desired final destination of the heterologous polypeptide in the animal. For example, in a preferred embodiment for the expression of heterologous polypeptides in mammal's milk, preferred host embryos are from animals which are already bred for large volume milk production, e.g., cows, sheep, goats and pigs.

There are standard procedures for introducing the DNA of the expression construct into animal cells. Commonly used transfection methods include electroporation [H. Potter et al., "Enhancer-Dependent Expression Of Human Kappa Immunoglobulin Genes Introduced Into Mouse Pre-B Lymhocytes By Electroporation", Proc. Natl. Acad. Sci. USA, 81(22), pp. 7161-65 (1984); G. Urlaub et al., "Isolation Of Chinese Hamster Cell Mutants Deficient In Dihydrofolate Reductase Activity", Proc. Natl. Acad. Sci. USA, 77(7), pp. 4216-4200 (1980)], protoplast fusion [R.M. Sandri-Goldin et al., Mol. Cell. Biol., 1, pp. 743-52 (1981)],

calcium phosphate coprecipitation [F.L. Graham and A.J. van der Eb, "A New Technique For The Assay Of Infectivity Of Human Adenovirus 5 DNA", Virology, 52(2), pp. 456-67 (1973); A.D. Miller et al., "c-fos Protein Can Induce Cellular Transformation: A Novel Mechanism Of Activation Of A Cellular Oncogene", Cell, 36(1), pp. 51-60 (1981)] and DEAE-dextran sulfate mediated protocols. In addition, many variations of the DEAE-dextran sulfate and calcium phosphate methods exist [C. Queen and D. Baltimore, "Immunoglobulin Gene Transcription Is Activated By Downstream Sequence Elements", Cell, 33(3), pp. 741-48 (1983); C.M. Gorman et al., "Recombinant Genomes Which Express Chloramphenicol Acetyltransferase In Mammalian Cells", Mol. Cell. Biol., 2(9), pp. 1044-11 (1982); R.S. McIvor et al., "Expression Of A cDNA Sequence Encoding Human Purine Nucleoside Phosphorylase In Rodent And Human Cells", Mol. Cell. Biol., 5(6), pp. 1349-57 (1985)] which may offer certain advantages. For example, calcium phosphate coprecipitation procedures are particularly effective with mammalian cells, including CHO cells.

A selectable marker is usually cointroduced with the island of expression construct into mammalian cells as a separate piece of DNA so that those cells which incorporate the expression construct can be readily isolated. Useful selectable markers include dihydrofolate reductase, metallothionein, neo, gpt, and hisD among others. The selected cells are then tested for expression of the heterologous protein.

There are also standard techniques for introducing the expression construct into the genome of a mammalian embryo. One technique for transgenically altering a mammal is to microinject the island of expression construct into the pronucleus of the

fertilized mammalian eggs to cause one or more copies of the construct to be integrated into the genome and retained in the cells of the developing mammals. Briefly, microinjection involves isolating fertilized ova, visualizing the pronucleus and then injecting the DNA into the pronucleus by holding the ova with a blunt holding pipette (approximately 50 μm in diameter) and using a sharply pointed pipet (approximately 1.5 μm in diameter) to inject buffer containing DNA into the pronucleus. See, for example, D. Kraemer et al., "Gene Transfer Into Pronuclei Of Cattle And Sheep Zygotes", Genetic Manipulation of the Early Mammalian Embryo, pp. 221-27, Cold Spring Harbor Laboratory (1985); R.E. Hammer et al., "Production Of Transgenic Rabbits, Sheep And Pigs By Microinjection", Nature, 315, pp. 680-83 (1985); and J.W. Gordon and F.H. Ruddle, "Gene Transfer Into Mouse Embryos: Production Of Transgenic Mice By Pronuclear Injection", Methods in Embryology, 101, pp. 411-33 (1983).

Microinjection is preferably carried out on an embryo at the one-cell stage, to maximize both the chances that the injected DNA will be incorporated into all cells of the animal and that the DNA will also be incorporated into the germ cells so that the animal's offspring will be transgenic as well. Usually, at least 40% of the mammals developing from the injected eggs contain at least one copy of the cloned construct in somatic tissues and these "transgenic mammals" usually transmit the gene through the germ line to the next generation. DNA isolated from the tissue of the resulting transgenic mammal may be tested for the presence of the island of expression by Southern blot analysis. If one or more copies of the island of expression remains stably integrated into the genome of such transgenic mammals, it is possible to establish

permanent transgenic mammal lines carrying the island of expression construct.

The offspring of transgenically altered mammals may be assayed after birth for the incorporation of the island of expression construct into the genome. Preferably, this assay is accomplished by Southern hybridization of chromosomal material from the progeny using a probe corresponding to a portion of the heterologous polypeptide coding sequence. Those mammalian progeny found to contain at least one copy of the construct in their genome are grown to maturity. In a preferred embodiment of this invention, the female species of these progeny will produce the desired heterologous polypeptide in or along with their milk. Alternatively, the transgenic mammals may be bred to produce other transgenic progeny useful in producing the desired heterologous polypeptides.

EXAMPLES

EXAMPLE 1 - CONSTRUCTION OF THE BOVINE ALPHA S-1 CASEIN ISLAND OF EXPRESSION

One example of this technology is to utilize the island of expression construct to produce a heterologous protein in a specific tissue or organ system of an intact animal. In this case we directed high level expression of a heterologous protein in the mammary gland of a mammal.

The gene construct described here contains an "island of expression" in which large 5' and 3' flanking regions of genomic sequence from the bovine alpha casein gene direct expression of the genomic clone of human urokinase. The 5' flanking region consists of 21 kb of upstream alpha casein sequences, including the first non-coding exon and the non-coding

portion of the second exon. The 9 kb 3' flanking region consists of the exons encoding the COOH-terminal half of alpha casein, the polyadenylation signal, and 2 kb of further downstream flanking sequences.

We cloned the bovine alpha S-1 casein gene (CAS) from a cosmid library of calf thymus DNA in the cosmid vector HC79 (from Boehringer Mannheim) as described by B. Hohn and J. Collins, "A Small Cosmid For Efficient Cloning Of Large DNA Fragments", Gene, 11(3-4), pp. 291-98 (1980). The thymus was obtained from a slaughterhouse and the DNA isolated by standard techniques well known in the art (T. Maniatis et al., Molecular Cloning: A Laboratory Manual at page 271, Cold Spring Harbor Laboratory 1982)). We constructed the cosmid library using standard techniques (F. Grosveld et al., "Isolation Of Beta - Globin - Related Genes From A Human Cosmid Library", Gene, 13(3), pp. 227-31 (1981)). We partially digested the calf thymus DNA with Sau3A (New England Bio Labs) and ran it on a NaCl gradient (1M to 5M) to enrich for 30 to 40 kb fragments. The partially digested DNA fragments were then ligated into the BamHI digested HC79 cosmid vector, followed by in vitro packaging by lambda extracts (Amersham Corporation, Arlington Heights, IL) according to the manufacturer's instructions. The in vitro packaged material was then used to transfect the E.coli K-12 strain HB101. Clones incorporating this vector were selected by growth on LB plates containing 50 μg/ml of Ampicillin (Sigma Chemical Co., St. Louis, MO).

We screened the resulting library using a 45 base pair oligonucleotide probe, CAS-1. This CAS-1 sequence, 5'-ATGGCTTGATCTTCAGTTGATTCACTCCCAATATCCTTGC TCAG-3', was synthesized based upon a partial cDNA sequence of alpha S-1 casein described by I.M. Willis

et al., "Construction And Identifiction By Partial Nucleotide Sequence Analysis Of Bovine Casein And Beta-Lactoglobulin cDNA Clones", DNA, 1(4), pp. 375-86 (1982). This sequence corresponds to amino acids 20-35 of mature bovine casein. As a result of this screening, we isolated three clones containing cosmids (C9, D4 and E1).

The 5' and 3' flanking sequences were obtained from cosmid clones, E1 and C9. Restriction mapping and Southern blot analysis (E. Southern, "Detection Of Specific Sequences Among DNA Fragments Separated By Gel Electrophoresis", J. Mol. Biol., 98 (3), pp. 503-517 (1975)) using oligonucleotide probes corresponding to known sequenced regions of the casein cDNA (A.F. Stewart et al., "Nucleotide Sequences Of Bovine Alpha S1- And Kappa-Casein cDNAs", Nucleic Acids Res., 12(9), pp. 3895-3907 (1984); M. Nagao et al., "Isolation And Sequence Analysis Of Bovine Alpha S1-Casein cDNA Clone", Agric. Biol. Chem., 48(6), pp. 1663-67 (1984)) established that cosmids D4 and E1 contained part of the casein structural gene (DNA sequence coding for the casein protein) and 21 kb of upstream or 5' flanking sequences. The C9 cosmid contained part of the casein structural gene and extended to 7 kb downstream of the polyadenylation sequence. We sequenced the cosmids E1 and D4 in the region corresponding to the transcriptional start of the casein structural sequence and determined that the sequence corresponded to that of a published sequence of the same region. (L.Y. Yu-Lee et al., "Evolution Of The Casein Multigene Family: Conserved Sequences In The 5' Flanking And Exon Regions", Nucleic Acid Res., 14(4), pp. 1883-1902 (1986)).

The construction of this island of expression in this invention is depicted in Figure 2. From the C9

cosmid we subcloned the 9 kb BamHI fragment which begins at a BamHI site within the intron following amino acid # 98 of alpha casein and continues to another Bam site located 2 kb downstream of the polyadenylation signal of alpha casein. This fragment is labelled as "C-term" in the Figure 2. This 9 kb fragment was cloned into BamHI-cut pUC19 to yield pCAS947. The downstream BamHI site was converted to a SalI site by partial digestion of pCAS947 with BamHI and subsequent ligation with a SalI linker, CAS 10, having the sequence, 5'-GATCGTCGAC-3'. The resulting plasmid was termed pCAS1238. This 9 kb BamHI-SalI fragment was used as the 3' flanking sequence of the "island". It contains the 3' untranslated region and 3' expression control sequences and a portion of the structural gene from alpha S-1 casein.

The next step was to design the 5' flanking region. The region containing the transcriptional start, a non-coding exon and a second exon, part of which was also non-coding, was subcloned. A 4 kb SmaI/BamHI fragment from cosmid E1 was isolated and subcloned into BamHI/SmaI-cut pUC19 to yield pCAS1176. The plasmid was cut with BglII, to remove the coding part of the second exon, and then the BglII site was converted to a BamHI site by ligation to a CAS 12 linker having the sequence, 5'-GATCTTGGATCCAA-3'. The resulting plasmid, pCAS1181, was then digested with SmaI and BamHI to remove the 3 kb piece of cosmid E1 DNA. The fragment was isolated, ligated to the 9 kb BamHI-SalI fragment from pCAS1238, and inserted into the SmaI/SalI digested pUC19 to yield pCAS1276.

The resulting construct links the transcriptional start site to the downstream genomic sequence with a unique BamHI cloning site in between, into which the heterologous polypeptide encoding

sequence can be inserted. Since the final constructs will have several other BamHI sites in the genomic sequences, the heterologous polypeptide encoding sequence cloning site was changed to both an XhoI site and a NotI site by the addition of a linker, CAS 30, having the sequence, 5'-GATCTCGAGCGCGGCCGCGCT-3'. The resulting vector, pCAS1277, contains XhoI and NotI sites as cloning sites in between the transcriptional start of alpha casein and the C-terminal genomic portion of alpha casein.

The transcriptional start and C-term regions from pCAS1277 were then used to replace the corresponding portions of the alpha casein genomic sequence found in the cosmid E1. Since the construct is 39 kb in length, cosmid technology was used to manipulate the plasmids. The original E1 cosmid was partially digested with XmaI, followed by digestion to completion with SalI to remove the 3'- most portion of the alpha casein gene contained in that cosmid. The SmaI and XmaI enzymes have the same recognition site, except that XmaI leaves a 5' overhang whereas SmaI leaves a blunt end. The 12 kb XmaI-SalI fragment from pCAS1277 was then inserted into the XmaI/SalI-cut cosmid to replace the removed portion.

The ligated products were subjected to *in vitro* packaging using an *in vitro* packaging kit (Amersham Corporation) and the packaged DNA was used to transfect *E.coli* DH5 cells, followed by selection on LB plates containing 50 μg/ml of ampicillin (Sigma Chemical Co.). The plasmids from ampicillin-resistant colonies were screened using oligonucleotide probes specific for the 3' end of casein. We identified and characterized plasmids which contain 21 kb upstream of the transcriptional start and the XhoI/NotI cloning site along with the genomic 3' end of the casein gene.

One of these plasmids, CAS1288, was then used to express the heterologous DNA sequence.

The genomic clone of human urokinase was isolated from a genomic library using published sequences as probes. A. Riccio et al., supra. From the published sequence, it can be seen that there is an ApaI site upstream of the translational start of the gene and also downstream of the polyA transcriptional signal. Oligonucleotide adapters (URO 8, having the sequence 5'-CGTCGACG-3', and URO 9, having the sequence 5'-GTACCGTCGACGGGCC-3') were used to add SalI sites to these two flanking ApaI sites. This allowed the genomic clone to be placed downstream of the SV40 early promoter in an animal cell expression vector so that we could test for expression prior to insertion in the alpha casein island of expression. The resulting plasmid, pUK0409, directed expression of authentic human urokinase in transfected tissue culture cells. We therefore knew that the genomic clone was functional. The next step was to put the urokinase genomic clone into the XhoI cloning site of CAS1288. These steps are depicted in Figure 3.

The urokinase genomic clone was isolated as an 8 kb SalI fragment from pUK0409. The SalI overhanging ends are capable of ligating into the XhoI cloning site found in CAS1288. There is, however, another XhoI site in the 21 kb upstream region of alpha casein. We therefore carried out partial XhoI digestions, followed by ligation with the isolated SalI urokinase fragment (see Figure 3). Plasmids were isolated from colonies and screened for the presence and orientation of the urokinase DNA sequence. One of these plasmids, CAS1295, contained the uorkinase gene in the correct orientation as determined by restriction analysis. This plasmid contains in a 5'- to -3'

orientation, the 21 kb upstream region, the first non-coding exon and intron sequences of casein, the genomic sequence coding for urokinase, the 9 kb 3' genomic alpha casein region.

EXAMPLE 2 - TRANSGENIC INCORPORATION OF THE "ISLAND OF EXPRESSION" CONSTRUCT INTO MICE

In order to carry out transgenic experiments, the prokaryotic vector sequences present in CAS1295 were removed before injection into embryos. This was accomplished by digesting CAS1295 with ClaI and SalI, followed by gel electrophoresis in 1% agarose TBE (see Maniatis et al, supra). The 41 kb fragment corresponding to the eukaryotic sequences of the island of expression construct was cut out of the gel and the DNA isolated by electroelution. The DNA was then centrifuged overnight in an equilibrium CsCl gradient. We removed the DNA band from the gradient and dialyzed extensively against TNE buffer (5 mM Tris, pH 7.4, 5 mM NaCl and 0.1 mM EDTA, pH 8).

The procedure for transgenic incorporation of the desired genetic information into the developing mouse embryo is established in the art. We followed techniques set forth in B. Hogan et al., Manipulating The Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory (1986). We used an F1 generation (Sloan Kettering) cross between C57B1 and CB6 mice (Jackson Laboratories). Six week old females were superovulated by injection of Gestile (pregnant mare serum) followed by human chorionic gonadotropin two days later. The treated females were bred with C57B1 stud males 24 hours later. The preimplantation fertilized embryos were removed within 12 hours following mating for microinjection with DNA and implantation into pseudopregnant females.

After isolating the embryo, we first digested away the cumulus cells surrounding the egg with hyaluronidase. The island of expression construct was then injected into the pronucleus of the embryo until it swelled 30% to 50% in size. We then implanted the injected embryos into the oviducts of pseudopregnant F1 females. DNA from the tails of the resulting live offspring was probed with nick translated CAS1295 DNA to identify those animals which carried the island of expression contruct. Three transgenic animals were identified. These animals were mated and the progeny tested for the presence of the island of expression construct as described supra.

One of the transgenic lines, which carried 2-3 copies of the island of expression construct, passed the genetic material in a Mendelian manner. The females of this transgenic line, which carry the CAS1295 insert, all produce human urokinase in their milk at about 1 mg/ml, as determined by enzymatic assay. The urokinase is inhibited by the monoclonal antibody #394, specific for human urokinase (Americana Diagnostica, Inc., New York, NY).

The other two transgenic lines carried 20-50 copies of the construct but failed to pass the DNA to the next generation of mice. We believe that the inability of the high copy number lines to pass the genes is due to the high basal level of the urokinase during embryogenesis. Urokinase is normally expressed in fetal tissue (embryonic stem cells) and may function in development. The low basal level of urokinase expression from the casein expression control sequences would not interfere with development in those embryos inheriting two copies of the gene. However, if expression is dependent upon copy number, those lines

which have 20-50 copies would have 20-50 fold higher basal level and would therefore express enough urokinase to interfere with proper development. These results indicate that the level of urokinase expressed is copy number dependent.

EXAMPLE 3 - TRANSFECTION OF THE ISLAND OF EXPRESSION CONSTRUCT INTO ANIMAL CELLS

The island of expression construct and the the selectable marker pSV2-DHFR (available from the American Type Culture Collection (ATCC 37146)) which codes for the production of dihydrofolate reductase in mammalian cells, are cointroduced into DHFR$^-$ CHO cells by electroporation. This technique is chosen for its ability to produce host cells characterized by stably integrated foreign DNA at high copy numbers. European Patent Application 0 343 783 fully describes this technique and is incorporated herein by reference.

Prior to electroporation, the pSV2-DHFR plasmid is linearized by digestion overnight at 37°C with AatII. The island of expression sequences are isolated from the vector sequences by cutting with restriction enzymes as described in Example 2, followed by gel electrophoresis to allow separation and purification (Maniatis et al., supra). Salmon sperm DNA (200 µg), previously sonicated to 300-1000 bp fragments, is added to a mixture containing 200 µg of the linearized pSV2-DHFR and 0.5 mg/ml of the island of expression construct. To precipitate the mixture of DNAs, NaCl is added to a final concentration of 0.1 M. Next, 2.5 volumes of ethanol are added and the mixture is incubated for ten minutes on dry ice. After a ten minute centrifugation at 4°C, the ethanol is aspirated and the DNA pellet is air-dried for 15 minutes in a tissue culture hood. The DNA pellet is then resuspended in 800 µl of 1X HeBS (20 mM Hepes/NaOH, pH

7.05; 137 mM NaCl; 5 mM KCl; 0.7 mM Na_2HPO_4; 6mM dextrose) for at least two hours prior to electroporation.

Approximately 2 x 10^7 $DHFR^-$ CHO cells (subcloned from the clone designated CHO-DUKX-B1 of Urlaub and Chasis, "Isolation Of Chinese Hamster Cell Mutants Deficient In Dihydrofolate Reductase Activity", Proc. Natl. Acas. Sci. UAS, 77, pp. 4216-20 (1980)) are used for each electroporation. The $DHFR^-$ CHO cells are passaged on the day prior to electoporation and are approximately 50% confluent on 10 cm plates at the time of harvesting for electroporation. The $DHFR^-$ CHO cells are detached from the plates by trypsin treatment and the trypsin subsequently inactivated by the addition of 8.0 ml α^- medium (MEM alpha supplemented with ribonucleotides and deoxyribonucleotides (10 mg/L each of adenosine, cytidine, guanosine, uridine, 2'-deoxyadenosine, 2'-deoxyguanosine and 2'-deoxy-thymidine; 11 mg/L of 2'-deoxycytidine hydrochloride) (Gibco Laboratories, Grand Island, NY), 10% fetal bovine serum (Hazelton, Lenexa, KS) and 4 mM glutamine (M.A. Bioproducts, Walkersville, MD)) per plate. The cells detached from the plates are then collected and centrifuged at 1000 rpm for 4 minutes. The majority of the medium is aspirated off the cell pellet and the cells resuspended in the remaining residual media by flicking the tube.

The island of expression, pSV2-DHFR and salmon sperm DNA, suspended in 800 µl 1X HeBS, are then added to the $DHFR^-$ CHO cell suspension. The resulting mixture is immediately transferred to an electroporation cuvette. The capacitor of the electroporation apparatus is set at 960 µF and the voltage set at 300V. A single pulse, lasting approximately 10 milliseconds, is delivered to the

contents of the cuvette at room temperature. The cells are then incubated for 8-10 minutes at room temperature and then transferred to a 15 ml tube containing 14 ml of α^+ medium. The cells are centrifuged as above. After aspirating the medium, the wet cell pellet is resuspended by flicking the tube and fresh α^+ medium is added. The suspended cells are then seeded into culture plates in non-selective medium for 2 days to allow them to recover from electroporation and express the selective gene. Approximately 20-30% of the viable CHO cells are expected to incorporate the island of expression/pSV2-DHFR and thus survive the selection process. Therefore, approximately 1×10^7 total cells per 10 cm plate are seeded and cultured in a 37°C, 5.5% CO_2 incubator.

After a recovery period of two days, the cells are removed from the culture plates by trypsin treatment as described above, counted and seeded into six 10 cm plates at a density of about 1×10^6 cells per plate, in α^- medium (Sigma Chemical Co.). The cells containing the island of expression and pSV2-DHFR are selected after a 4 day incubation in the α^- media. The selected cells are then tested for expression of urokinase by standard techniques, e.g, a commercially available colorometric test, Spectrozyme UK (Americana Diagnostica, Inc.)

Several clones that have various levels of expression of urokinase are selected. DNA and RNA are isolated from these clones and Northern and Southern analysis is carried out to determine transcription level and copy number of the island of expression construct. This analysis reveals whether expression of the uorkinase message is a function of the copy number and independent of the site of integration of the integrated construct.

A construct according to this invention containing plasmid CAS1288 is exemplified by a culture deposited at In Vitro International, Inc. in Linthicum, Maryland, on February 1, 1990 and there identified as CAS1288 wherein the plasmid CAS1288 is in E.coli DH5. It has been assigned accession number IVI 10232.

A second construct according to this invention containing plasmid CAS1295 is exemplified by a culture deposited at In Vitro International, Inc. in Linthicum, Maryland, on February 1, 1990 and there identified as CAS1295 wherein the plasmid CAS1295 is in E.coli DH5. It has been assigned accession number IVI 10231.

While we have hereinbefore presented a number of embodiments of our invention, it is apparent that our basic construction may be altered to provide other embodiments which utilize the processes and compositions of this invention. Therefore, it will be appreciated that the scope of this invention is to be defined by the claims appended hereto, rather than the specific embodiments which have been presented hereinbefore by way of example.

We claim:

1. A process for producing a high the level of a desired heterologous polypeptide in a host, the process comprising the steps of:

a) integrating at least one island of expression into the genome of said host, wherein said island of expression comprises, in the 5' to 3' direction, a 5' flanking region, a heterologous polypeptide encoding sequence and a 3' flanking region; said 5' flanking region comprising 5' expression control sequences, operatively linked to said heterologous polypeptide encoding sequence and a 5' untranslated region; said 3' flanking region comprising, a 3' untranslated region, and 3' expression control sequences, operatively linked to said heterologous polypeptide encoding sequence; and the 5' and 3' flanking regions of said islands of expression being of sufficient size and structure effective to render the level of production of the desired heterologous polypeptide substantially dependent on the copy number of the island of expression integrated into the host genome and substantially independent of the position of integration of the island of expression in the host genome; and

b) culturing said host under conditions which allow said desired heterologous polypeptide to be expressed.

2. The process according to claim 1 wherein said heterologous polypeptide encoding sequence comprises a functional signal sequence coding region.

3. The process according to claim 2 wherein the signal sequence coding region is derived from a milk specific protein gene.

4. The process according to claim 3 wherein the milk specific protein gene is casein.

5. The process according to any one of claims 2 to 4 wherein the host is a lactating mammal selected from the group consisting of mice, cows, sheep, goats and pigs and the 5' and 3' flanking sequences are derived from a milk specific protein gene.

6. The process according to claim 1 wherein the heterologous polypeptide encoding sequence is selected from sequences encoding a polypeptide selected from the group consisting of: tPA, urokinase, Mullerian Inhibiting Substance, interferons, coagulation factors VIII and IX, animal growth hormones, insulin, interleukins, immunoglobulins and lipocortins.

7. An island of expression DNA sequence comprising, in the 5' to 3' direction, a 5' flanking region, a heterologous polypeptide encoding sequence and a 3' flanking region; the 5' flanking region comprising 5' expression control sequences operatively linked to the heterologous polypeptide encoding sequence and a 5' untranslated region; the 3' flanking region comprising a 3' untranslated region, and 3' expression control sequences, operatively linked to the heterologous polypeptide encoding region; whereinupon the integration of the island of expression into the genome of a host, the 5' and 3' flanking regions of the island of expression are of sufficient size and structure effective to render a level of production of a polypeptide encoded by the heterologous polypeptide encoding sequence substantially dependent on the copy

number of the island of expression in the host genome and substantially independent of the position of integration of the island of expression in the host genome.

8. The island of expression according to claim 7 wherein the heterologous polypeptide encoding sequence comprises a functional signal sequence coding region.

9. The island of expression according to claim 7 wherein the 5' and 3' flanking sequences are derived from a milk specific protein gene.

10. The island of expression according to claim 8 wherein the signal sequence coding region is derived from a milk specific protein gene.

11. The island of expression according to claim 9 wherein the milk specific protein gene is casein.

12. The island of expression according to claim 7 wherein the heterologous polypeptide encoding sequence is selected from sequences encoding a polypeptide selected from the group consisting of: tPA, urokinase, Mullerian Inhibiting Substance, interferons, coagulation factors VIII and IX, animal growth hormones, insulin, interleukins, immunoglobulins and lipocortins.

13. A transformed host characterized by a genome comprising an integrated island of expression, said island of expression according to any one of claims 7 to 12.

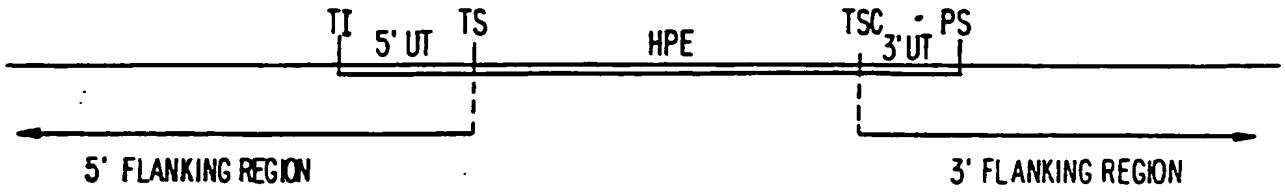
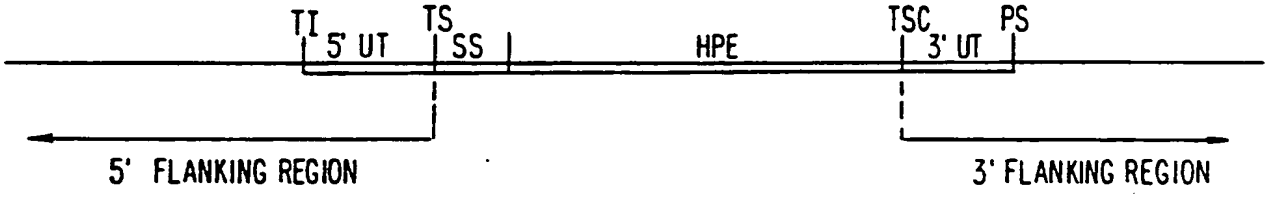


FIG. 1a



TI= TRANSCRIPTION INITIATION SITE
UT= UNTRANSLATED REGION
TS= TRANSLATIONAL START SIGNAL
SS= SIGNAL SEQUENCE ENCODING REGION
HPE= HETEROLOGOUS POLYPEPTIDE ENCODING REGION
TSC= TRANSLATION STOP CODON
PS= POLYADENYLATION SIGNAL

FIG. 1b

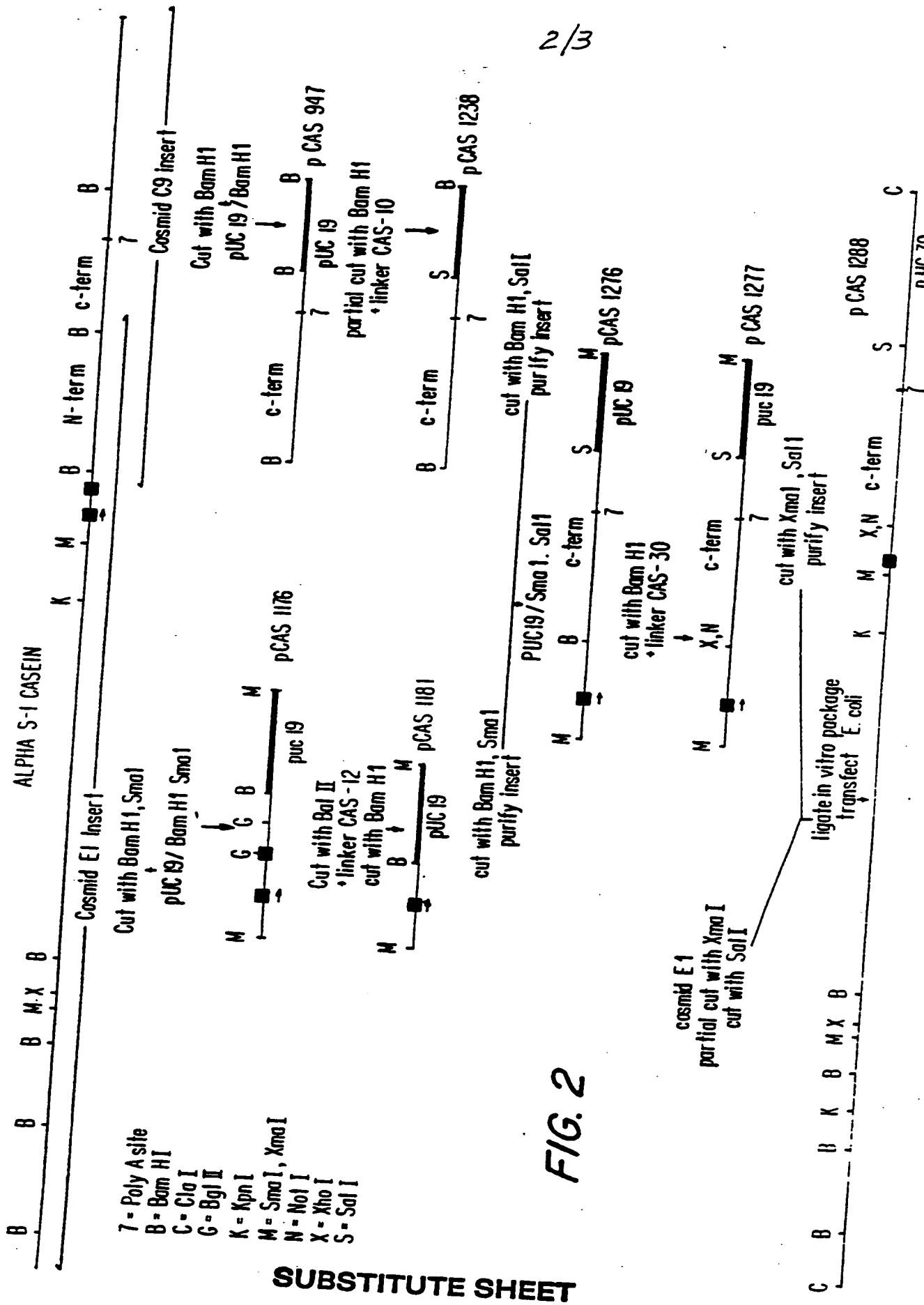

ALPHA S-1 CASEIN
Cosmid E1 insert
Cosmid C9 insert
Cut with Bam H1, Sma I
pUC 19/ Bam H1 Sma I
Cut with Bam H1
pUC 19 / Bam H1
partial cut with Bam H1
+ linker CAS-10
pCAS 1176
puc 19
Cut with Bal II
+ linker CAS-12
cut with Bam H1
pUC 19
pCAS 1181
cut with Bam H1, Sma I
purify insert
PUC19/ Sma I. Sal I
cut with Bam H1, Sal I
purify insert
pCAS 1276
pUC 19
cut with Bam H1
+ linker CAS-30
pCAS 1277
puc 19
cut with Xma I, Sal I
purify insert
cosmid E 1
partial cut with Xma I
cut with Sal I
ligate in vitro package
transfect E. coli
pCAS 1288
pUC 79
7 = Poly A site
B = Bam HI
C = Cla I
G = Bgl II
K = Kpn I
M = Sma I, Xma I
N = Not I
X = Xho I
S = Sal I

FIG. 2

FIG. 3a

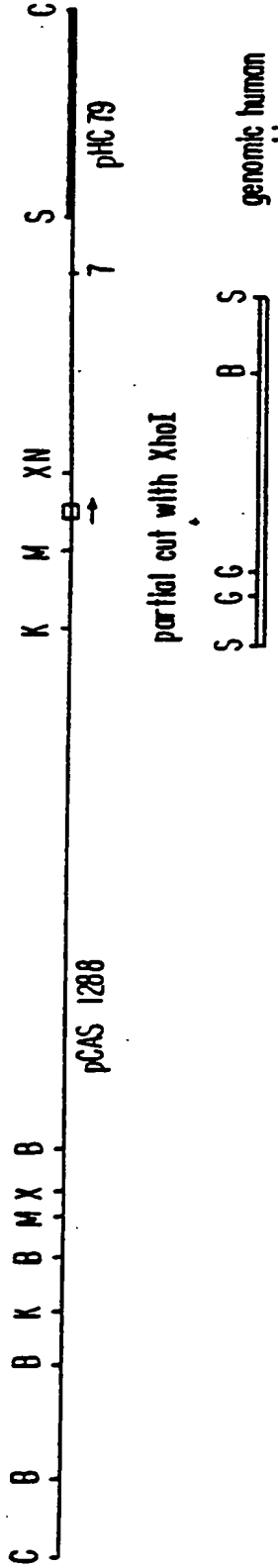
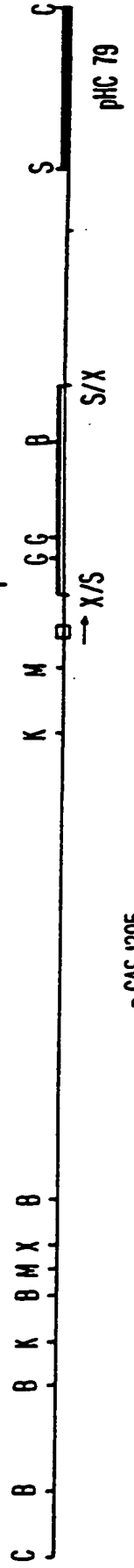

C B B K B M X B
pCAS 1288
K M XN
S G G
partial cut with XhoI
B S
genomic human urokinase
in vitro package
transfect E. coli
S 7 pHC 79
C

FIG. 3b


C B B K B M X B
pCAS 1295
K M G G
X/S
B S/X
S pHC 79
C

ANNEX M3

International Application No: PCT/

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 31, line 11 of the description [1]

A. IDENTIFICATION OF DEPOSIT [2] CAS1295 in E. coli DH5

Further deposits are identified on an additional sheet ☐ [3]

Name of depositary institution [4]

IVI International, Inc.

Address of depositary institution (including postal code and country) [5]

611 P. Hammonds Ferry Road
Linthicum, Maryland 21090 (USA)

Date of deposit [6]	Accession Number [7]
1 February 1990	IVI 10232

B. ADDITIONAL INDICATIONS [8] (leave blank if not applicable). This information is continued on a separate attached sheet ☐

In respect of those designations in which a European patent is sought samples of the deposited microorganisms will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE [9] (if the indications are not for all designated States)

EPO

D. SEPARATE FURNISHING OF INDICATIONS [10] (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later [11] (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

(Authorized Officer)

☒ The date of receipt (from the applicant) by the International Bureau [12]

was 09 AUGUST 1991
(09.08.91)

(Authorized Officer)

Form PCT/RO/134 (January 1981)

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 91/01222

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[5]: C 12 N 15/00, 15/85, 15/90 15/62 // 15/58

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[5]	C 12 N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	WO, A, 8901517 (F. GROSVELD) 23 February 1989 see the whole document, in particular: page 7, line 17 - page 8, line 8; page 13, lines 9-12	1,7,13
Y	---	2-6,8-12
Y	WO, A, 8810118 (BIOGEN N.V.) 29 December 1988 see claims 3,4,6; figure 1 ---	2-6,8-12
P,X	BIO/ Technology, Volume 8, May 1990, (New York, US), H. Meade et al.: "Bovine $alpha_{s1}$-casein gene sequences direct high level expression of active human urokinase in mouse milk", pages 443-446 see the whole document ---	1-13

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
15th May 1991	28 JUN 1991
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	Miss I. TAZELAAR

Form PCT/ISA/210 (second sheet) (January 1985)

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9101222 SA 45412

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 21/06/91
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A- 8901517	23-02-89	AU-A-	2137388	09-03-89
		EP-A-	0332667	20-09-89
		JP-T-	2500802	22-03-90
WO-A- 8810118	29-12-88	US-A-	4873316	10-10-89
		EP-A-	0347431	27-12-89
		JP-T-	2500798	22-03-90